

# Latent Change Score Modeling as a Method for Analyzing the Antidepressant Effect of a Psychosocial Intervention in Alzheimer's Disease

Katja Werheid<sup>a</sup> Ylva Köhncke<sup>a, d</sup> Matthias Ziegler<sup>b</sup> Alexander Kurz<sup>c</sup>

<sup>a</sup>Clinical Gerontopsychology, and <sup>b</sup>Psychological Diagnostics, Department of Psychology, Humboldt Universität zu Berlin, Berlin, and <sup>c</sup>Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; <sup>d</sup>Aging Research Center, Karolinska Institutet and Stockholm University, Stockholm, Sweden

## Key Words

Structural equation modeling · Dementia · Depression

## Abstract

**Background:** Developing and evaluating interventions for patients with age-associated disorders is a rising field in psychotherapy research. Its methodological challenges include the high between-subject variability and the wealth of influencing factors associated with longer lifetime. Latent change score modeling (LCSM), a technique based on structural equation modeling, may be well suited to analyzing longitudinal data sets obtained in clinical trials. Here, we used LCSM to evaluate the antidepressant effect of a combined cognitive behavioral/cognitive rehabilitation (CB/CR) intervention in Alzheimer's disease (AD). **Methods:** LCSM was applied to predict the change in depressive symptoms from baseline as an outcome of the CORDIAL study, a randomized controlled trial involving 201 patients with mild AD. The participants underwent either the CORDIAL CB/CR program or standard treatment. Using LCSM, the model best predicting changes in Geriatric Depression Scale scores was determined based on this data set. **Results:** The best fit was achieved by a model predicting a decline in depressive symptoms between before and after testing. Assignment to the intervention group

as well as female gender revealed significant effects in model fit indices, which remained stable at 6- and 12-month follow-up examinations. The pre-post effect was pronounced for patients with clinically relevant depressive symptoms at baseline. **Conclusions:** LCSM confirmed the antidepressant effect of the CORDIAL therapy program, which was limited to women. The effect was pronounced in patients with clinically relevant depressive symptoms at baseline. Methodologically, LCSM appears well suited to analyzing longitudinal data from clinical trials in aged populations, by accounting for the high between-subject variability and providing information on the differential indication of the probed intervention.

© 2015 S. Karger AG, Basel

## Introduction

The worldwide rise in life expectancy will exert an effect on research in psychotherapy [1]. There is an urgent need to evaluate existing interventions for the elderly and to develop or adapt interventions for patients with age-associated disorders [2]. However, research in this field faces major methodological challenges. As a well-established finding in lifespan research, between-subject vari-

ability increases with age [3]. Especially above the age of 70 years, the range of diversity in cognitive and somatic functioning, as well as in quality of life, is broader than in earlier life periods [4].

As a consequence of age-associated diversity, the impact of factors which influence the outcome of psychotherapy rises with age. Conventional statistical approaches such as general linear model (GLM)-based data analyses – for instance, *t* testing, ANOVAs, and ANCOVAs using pre-post differences in outcome measures between intervention and control groups – come along with a number of strong assumptions. Most importantly for clinical research, the assumption of normally distributed variables is likely to be violated when using clinical samples. Whereas such violations can be regarded as less critical [5], they become increasingly problematic if group variances are heterogeneous [6]. Even though remedies such as the Welch or Brown-Forsythe test have been introduced [7], it remains problematic that all of these approaches rely on approximating minimal variance within groups, which might, in older samples, imply a bias towards underestimating treatment effects.

ANCOVAs, which are often recommended as the method of choice when comparing two groups in a pre-post design [8], have been shown to result in disrupted significance levels if unequal regression slopes and unequal group sizes go along with nonnormality. Thus, in the case of clinical studies where the influence of the precondition on the postcondition is not equal across groups, ANCOVA results might be problematic. Recent clinical studies often use more than two measurement points, which further lowers the usability of ANCOVAs. This leaves repeated-measures ANOVA as another method of choice. However, it has been shown that violations due to group effects (e.g., patients all being in the same clinic) can substantially influence the results [9]. A further critical issue leading to false-positive results in repeated-measures ANOVAs is the violation of the sphericity assumption [10]. Sphericity means that the variance of all pairwise differences between measures is constant. This assumption is very likely to be violated if therapy effects are not uniform. Thus, common approaches to testing therapy effects across time are problematic for several reasons.

In addition to the likelihood of incorrectly estimating the significance of treatment effects, the GLM-based rationale bears the risk of limiting our knowledge regarding differential indication in clinical studies [11]. It leaves open the clinician's most relevant question in view of short-running health care resources, that is, whether psy-

chotherapy will be more or less effective for an older patient with specific demographic or medical characteristics. In other words, GLM-based approaches are restricted by the way differential therapy outcomes can be explored.

Latent change score modeling (LCSM) [12, 13], a structural equation modeling-based technique, may represent an alternative approach. Structural equation modeling-based analyses have been widely used in large longitudinal studies and aging research within the previous decade [14–16]. While this method has proven to be useful in epidemiological studies involving neurological patients [17], it is rarely applied in clinical trials. However, LCSM might be especially well suited to this purpose as it permits the identification of relationships between baseline levels and changes in individual trajectories. It also allows the researcher to determine the influence of preselected factors on change in the outcome score. Moreover, LCSM makes it possible to examine changes that follow upon each other. Finally, it is well suited to detecting complex patterns of covariances in larger sets of continuous variables. Moreover, modern statistical packages provide a number of different robust estimators dealing with nonnormality or ordinal data. Thus, in contrast to traditional GLM-based methods, LCSM not only is more robust but even relies on broad between-subject variability, a core feature of data obtained from the aged.

In the present study, LCSM was applied to data obtained from the CORDIAL study [18], a clinical trial investigating the effectiveness of a multicomponent intervention combining cognitive behavioral (CB) and cognitive rehabilitation (CR) approaches to Alzheimer's disease (AD). The intervention consisted of 12 weekly individual 1-hour sessions, organized in 4 thematic modules, flanked by an introductory and a closing module. The thematic therapy modules focused on using compensatory memory aids (module 2), establishing daily routines (module 3), enhancing pleasant activities (module 4), and self-supporting reminiscence (module 5). The study involved 201 patients from 5 outpatient units at university clinics and from 5 neuropsychiatric practices. All centers strictly followed the same protocol and adhered to the same criteria for subject enrolment and assessment as well as administration of interventions. Participants were assessed at baseline and after the intervention (at 0 and 3 months) as well as in 2 follow-ups 6 and 12 months thereafter. The primary outcome measure was the change in functional ability from baseline to after testing as measured by the Bayer Activities of Daily Living (B-ADL) scale. Secondary outcome measures were a change as described above in

Dementia Quality of Life scale (DEMQOL), Neuropsychiatric Inventory (NPI), a cognitive test battery, a patient satisfaction scale (ZUF-8), and the Geriatric Depression Scale (GDS). Further details of this study, not yet involving the 12-month follow-up, have been reported in the original paper [18]. Of importance for the present study, GLM-based comparisons of the intervention and control groups indicated that after the test and at the 6-month follow-up, depressive symptoms were significantly reduced in women but not in the total sample. Otherwise, there were no significant changes in the outcome measures.

Our current re-analysis of the data was motivated by two aspects of the study. First, as in most studies on psychosocial interventions in AD, the patient sample was not stratified for gender. The gender ratio was 1:1 instead of 1:2, which would correspond to the higher prevalence of AD in elderly women [19, 20]. Second, the degree of depressiveness as measured by the GDS at baseline greatly varied between 0 and 28 symptoms reported. For the reasons explained above, these two aspects gave rise to the hypothesis that the antidepressant effect of the intervention might have been underestimated by the GLM-based methodology. In the present study, we therefore re-analyzed the data of the CORDIAL study in order to examine the combined effects of group assignment and gender on depressiveness in early AD. An optimal model fit was intended for the total sample, and additionally for the subsample with clinically relevant depressive symptoms.

## Subjects and Methods

### Participants

All participants were community-dwelling patients diagnosed with mild dementia in AD (ICD-10 research criteria, MMSE score >20) at an average age of 73.7 years (SD = 7.9) who did not have any other psychiatric or physical disorders. They were randomly assigned to either an intervention group (IG) or a control group (CG) and assessed 4 times by blinded, trained raters: at baseline before the intervention ( $t_0$ ) and approximately 3 months later directly after the intervention ( $t_1$ ) as well as at 2 follow-ups ( $t_2$  and  $t_3$ ) 6 and 12 months after the intervention. The CG was slightly older (mean = 75.1 years, SD = 7.1,  $p = 0.016$ ) than the IG (mean = 72.4 years, SD = 8.5). The dropout rates were very favorable compared to clinical studies in general and those in AD specifically: 8 and 4% at  $t_1$ , 17 and 12% at  $t_2$ , and 24 and 21% at  $t_3$  for the IG and the CG, respectively. Dropout at  $t_1$  depended on an interaction of center, gender, and group ( $F = 2.60$ ,  $df = 9$ ,  $p < 0.01$ ); subsequent dropout was not associated with these variables. Dropout associated with center, gender, and group was handled in the main analyses. A detailed description of the intervention, participant characteristics, and study design is given in Kurz et al. [18].

### Assessment of Depressive Symptoms

The primary outcome for the present analyses was the GDS [21], German version [22], a 30-item self-report measure to be confirmed or rejected by the patient, yielding sum scores between 0 and 30. Scores of  $\geq 10$  are considered clinically relevant signs of depression [23, 24]; they were therefore selected as a criterion for depression as part of the BPSD (behavioral and psychological symptoms of dementia) syndrome in AD.

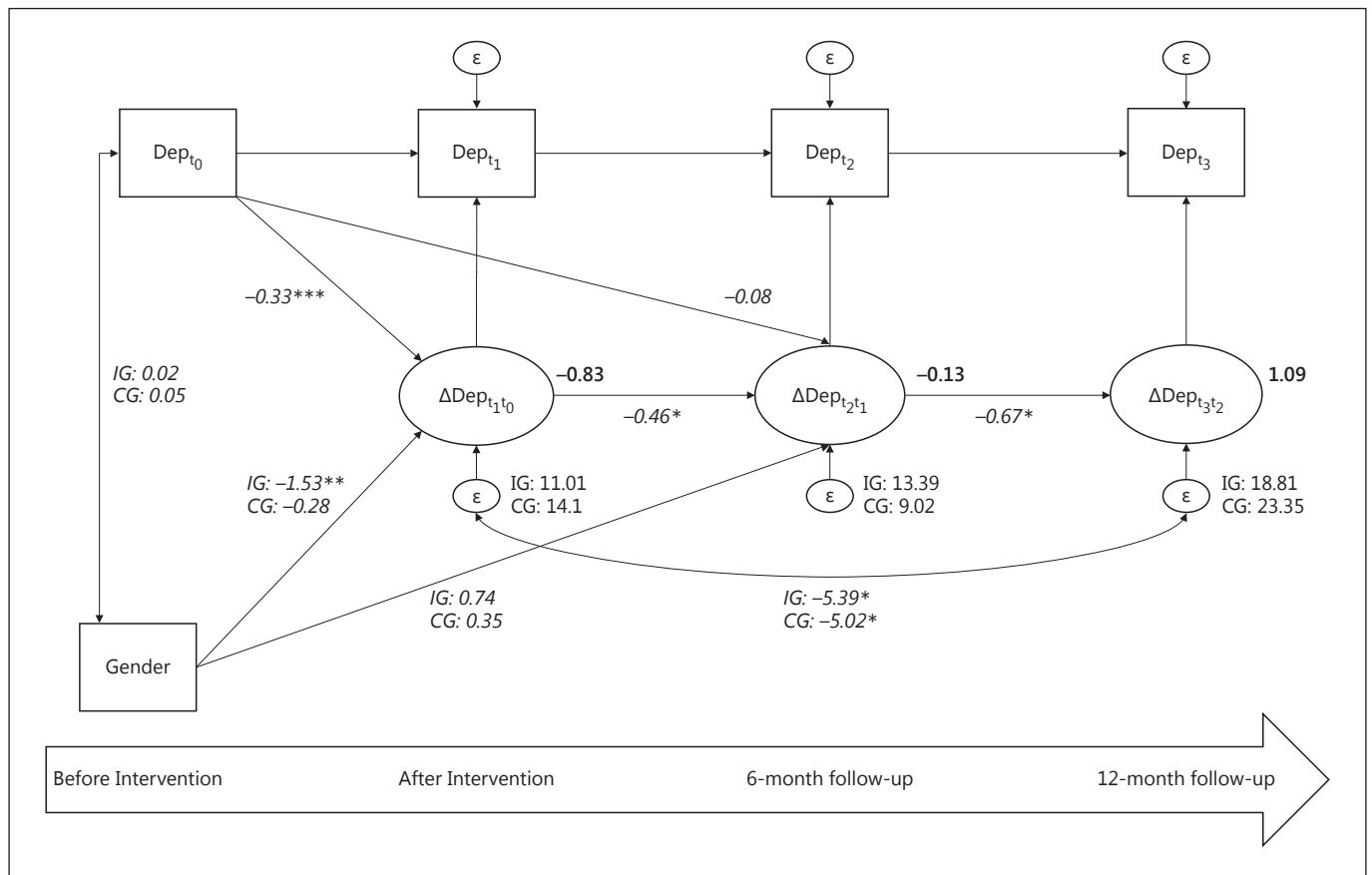
### Statistical Analyses

Analyses were based on the data on the intention-to-treat population. LCSM was applied [13] using Mplus version 5.2 [25]. In order to systematically test alternative assumptions, nested multiple group models (fig. 1) were compared. Aiming for a parsimonious yet well-fitting model, we systematically introduced constraints to the model just identified and tested for a resulting loss of model fit. One by one, we constrained each parameter to 0, and if that led to worse fit, we constrained it to be of the same value across groups. If this led to worse fit, we concluded there must be group differences in this parameter and allowed the model to estimate them [17, 26]. This way, we kept each parameter that, if constrained, resulted in a loss of model fit.

Loss of model fit was defined as a significant difference in likelihood as tested by the  $\chi^2$  statistic ( $\alpha$  level:  $p = 0.05$ ). As this test might be underpowered with small samples, we supplemented it with the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Whenever the  $\chi^2$  difference test indicated a significant worsening of model fit after adding a constraint to the model, we refrained from that constraint. When it was nonsignificant, we inspected the other fit indices mentioned above and refrained from a constraint whenever we simultaneously observed a decrement of  $>0.005$  in the CFI, a change of  $>0.010$  in the RMSEA, and a change of  $>0.025$  in the SRMR (for intercepts, a change  $>0.005$ ). Simulations have demonstrated that these cutoffs are sensitive to test measurement invariance across groups [27]. Thus, we iteratively tested whether constraining parameters led to a decrease in model fit to explore group differences. By doing so, we explicitly examined possible group differences in all modeled parameters. Finally, we combined all possible constraints in a parsimonious model. This final model then assumedly contained all parameters necessary to model group differences in the data. In a next step, we evaluated the estimated values for each parameter to assess the strengths of the associations or differences as well as their statistical significance. We achieved this by again constraining one parameter after the other to 0 or, if constraint to 0 was impossible, to an equal value across groups. The  $p$  value from the  $\chi^2$  difference test indicated the statistical significance of a parameter estimate.

The models were estimated using a robust estimator (MLR) implemented in Mplus 7 [25].  $\chi^2$  difference tests were corrected with the scaled formula by Satorra and Bentler [28]. All the estimated models included 'study center' as a cluster variable to account for clustered sampling.

Missing data due to dropout were dealt with by full information maximum likelihood estimation [29]. The main advantage of this algorithm is that it takes into account all information available from the variables included in the model when estimating parameters; that is, even cases with partly missing data are included without former imputation of the missing values. This was shown to



**Fig. 1.** Preferred model, unstandardized path coefficients, intercepts and covariances for the CG and the IG, respectively. Intercepts (in bold type) of the change scores ( $\Delta\text{Dep}$ ) represent change in depressive symptoms for a person with the average score of the respective group. Gender was coded binarily (men = 0, women = 1), so that the intercepts of the change scores represent change for an average male patient of the respective group, whereas the path coefficient gender – latent change score (in italics) adds to the intercept to depict change for an average female patient. When just

one coefficient is depicted, it is set equal in both groups. If no coefficients are depicted, the path is fixed to 1. Plain: residual variance; bold type: intercepts; italics: path coefficients, covariances. \*/\*\*/\* Significance according to the  $\chi^2$  difference test, testing this model against one in which the respective parameter is fixed. Residual variances were not tested. Fit indices:  $\chi^2$  value (df = 11) = 9.74; RMSEA = 0 (90% CI: 0–0.095); CFI = 1; SRMR = 0.062.  $\epsilon$  = latent variable.

yield less bias in the estimated population parameters than other widely used methods [30].

The respective influence of gender on the change scores was measured by testing the increment in explained variance when including gender into the model. Observed GDS scores were centered on the group-specific average score at  $t_0$ , so that the intercepts of the change scores represent change in depressive symptoms for a person from the average score of the respective group. Gender was coded binarily (men = 0, women = 1), so that the mean difference between average male and female patients is estimated by the path coefficient gender – latent change score. The intercepts of the change scores represent change for an average male patient of the respective group, whereas the path coefficient gender – latent change score adds to the intercept to depict change for an average female patient (see fig. 1).

For all t tests or ANOVAs, homogeneity of variances was tested with Levene's test, and, if significant, corrected values are reported. For ANOVAs, Bonferroni-corrected values are reported when appropriate. t tests and ANOVAs were conducted with SPSS 19 (IBM), and effect sizes were calculated using G\*Power [31]. Significance was tested using a 5%  $\alpha$  level (two-tailed).

## Results

### Model Fit for the Total Sample

At baseline, the number of depressive symptoms did neither differ between men and women ( $p = 0.96$ ) nor

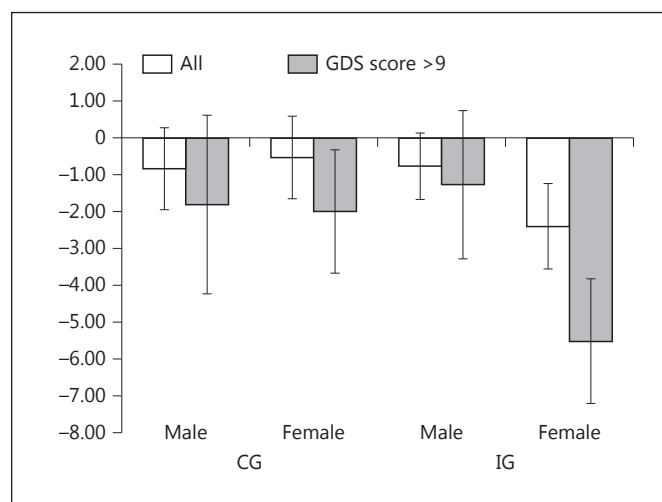


between groups ( $p = 0.33$ ). The most parsimonious model, henceforth defined as the preferred model (fig. 1), showed a good model fit [ $\chi^2$  value ( $df = 11$ ) = 9.74; RMSEA = 0 (90% CI: 0–0.095); CFI = 1; SRMR = 0.062]. After being pruned in the way described, the preferred model estimated GDS change over time, GDS baseline effects on GDS change, and gender effects on GDS change. Importantly, group differences were estimated in the effect of gender (group by gender interaction) on GDS change during the intervention ( $\Delta\text{Dep}_{t_1t_0}$ ) and during the first follow-up period ( $\Delta\text{Dep}_{t_2t_1}$ ). The preferred model included the following constraints: (a) to be 0, the influence of baseline GDS score and of gender on GDS score change in the second follow-up period ( $\Delta\text{Dep}_{t_3t_2}$ ), and (b) to be equal across groups, the influence of baseline GDS score on  $\Delta\text{Dep}_{t_1t_0}$  and  $\Delta\text{Dep}_{t_2t_1}$  as well as the intercepts of all  $\Delta\text{Dep}$  scores (fig. 1).

According to the preferred model, an averagely depressed male patient would show a nonsignificant decrease in symptoms by 0.83 GDS points ( $\beta_0 = -0.83$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 2.03$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.154$ ) during the intervention period irrespective of group assignment. In contrast, for female patients, the extent of decrease depended on group assignment (fig. 2). This was evident in a significant path coefficient from gender to  $\Delta\text{Dep}_{t_1t_0}$  in the IG ( $\beta_{\text{unstd}} = -1.52$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 6.47$ ,  $\Delta df = 1$ ,  $p_{(\Delta\chi^2)} = 0.011$ ) but not in the CG ( $\beta_{\text{unstd}} = 0.28$ ,  $\Delta\chi^2_{\text{S-B-scaled}} = 0.28$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.60$ ), meaning that being a female patient resulted in a 1.52 times higher GDS score after the intervention as compared to being a male patient. The difference between the groups in the gender effect was significant ( $\Delta\chi^2_{\text{S-B-scaled}} = 8.53$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.003$ ), indicating a group by gender interaction.

As a further result of the preferred model, in both groups, the number of symptoms at baseline affected the first change score ( $\beta_{\text{unstd}} = -0.33$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 94.21$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} < 0.001$ ); that is, 1 point higher on the GDS at baseline went along with a decrease of 0.33 GDS points during the intervention, in addition to the  $-0.83$ -point change intercept for male patients and the  $-0.83$ -point change intercept plus  $-1.52$ -point gender effect for female patients. In the CG, the model explained 17.9% of the total variance of the first change score; a model without gender explained 17.5%. In the IG, the model explained 22% of the total variance; a model without gender explained 17.6%. Both overall effect sizes can be considered as moderate to large [32].

Regarding the follow-up periods, the only significant group difference was to be found in the effect of gender on  $\Delta\text{Dep}_{t_2t_1}$ , that is, the first follow-up period. In the CG,



**Fig. 2.** Mean change in depressive symptoms during the intervention as predicted by the preferred model, according to gender and group. Error bars: 95% CIs. Note that the depicted means of the latent change factor scores can be directly interpreted for prognostic use. For example, the model predicts that the GDS score of a female patient ranging above cutoff at baseline would drop by 5.53 points if the intervention were applied, but only by 1.81 points if the intervention were not applied.

women showed slightly less of a decrease in depressive symptoms than men, as indicated by a nonsignificantly positive effect of gender on  $\Delta\text{Dep}_{t_2t_1}$  ( $\beta_{\text{unstd}} = 0.35$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 0.26$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.61$ ). In the IG, women showed slightly more of a decrease in depressive symptoms than men, as indicated by a nonsignificantly positive effect of gender on  $\Delta\text{Dep}_{t_2t_1}$  ( $\beta_{\text{unstd}} = -0.74$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 0.97$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.33$ ). Group differences in gender effects (group  $\times$  gender interaction) were also nonsignificant ( $\Delta\chi^2_{\text{S-B-scaled}} = 1.1$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.29$ ). The intercept of  $\Delta\text{Dep}_{t_2t_1}$  was nonsignificant in either group, indicating no further mean change for men irrespective of group ( $\beta_0 = -0.13$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 0.11$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.74$ ). The preferred model explained 26.6% of the total variance in  $\Delta\text{Dep}_{t_2t_1}$  (without gender: 25%) in the CG. In the IG, the model explained 15.2% of the variance in  $\Delta\text{Dep}_{t_2t_1}$  (without gender: 14.7%). Both overall effects can be regarded as moderate to large [32].  $\Delta\text{Dep}_{t_3t_2}$  had a positive intercept ( $\beta_0 = 1.1$ ;  $\Delta\chi^2_{\text{S-B-scaled}} = 3.46$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.06$ ); it was set to be equal across groups and genders, indicating a mean increase in depressive symptoms across groups and genders between 6 and 12 months after the intervention. In the CG, the model explained 9.3% of the variance in  $\Delta\text{Dep}_{t_3t_2}$  (without gender: 9.4%), and in the IG, it explained 16.9% of the variance (without gender: 16.5%),

indicating that previous GDS, but not gender, had a moderate effect on follow-up GDS change in both groups.

Our fourth step of analysis addressed the within-person stability of the GDS scores in the follow-up periods. Each  $\Delta$  negatively influenced the next  $\Delta$  (fig. 1), indicating fluctuations such as that high scores tended to drop more and vice versa. There were no significant group differences in these parameters, meaning that these fluctuations were similar in both groups.

#### *Model Fit for Clinically Depressed Patients*

A similar model, but reduced to 3 measurement occasions, was tested in a subsample of clinically depressed patients, defined by a GDS score of  $\geq 10$  [23]. At  $t_0$ , the depressive subsample differed from the nondepressive counterpart with respect to GDS score (mean = 14.2 vs. 5.5;  $t_{df=199} = -20.67$ ,  $p < 0.001$ , power = 1; effect size  $d = 2.8$ ) but not to age (mean = 73.8 vs. 73.7 years;  $t_{df=199} = 0.015$ ,  $p = 0.99$ , power = 0.05; effect size  $d = 0.002$ ) or years of education (mean = 12.3 vs. 12.7 years;  $t_{df=199} = 0.95$ ,  $p = 0.34$ , power = 0.12; effect size  $d = 0.11$ ). Importantly, being in the subsample was neither affected by group assignment nor by gender (ANOVAs:  $p > 0.71$ ). Within the subsample, GDS scores at  $t_0$  did not differ between men and women [difference between means ( $\text{diff}_{M-M}$ ) = 0.23,  $t_{df=77} = 0.29$ ,  $p = 0.77$ , power = 0.06; effect size  $d = 0.07$ ] or between groups ( $\text{diff}_{M-M} = 0.79$ ,  $t_{df=77} = 1.02$ ,  $p = 0.31$ , power = 0.17; effect size  $d = 0.23$ ). Due to the small sample size of the subsamples (CG:  $n = 41$ ; IG:  $n = 38$ ), reliable testing was only possible for a model with 3 time points.

The preferred model allowed for the following constraints. The ensuing parameters could be set equal across groups: intercepts of both  $\Delta\text{Dep}_{t_1t_0}$  and  $\Delta\text{Dep}_{t_2t_1}$  (GDS change in men), influence of baseline GDS score on  $\Delta\text{Dep}_{t_1t_0}$ , and influence of  $\Delta\text{Dep}_{t_1t_0}$  on  $\Delta\text{Dep}_{t_2t_1}$ . Effects of gender on  $\Delta\text{Dep}_{t_2t_1}$  could be set to 0. The preferred model fit the data well [ $\chi^2$  value ( $df = 6$ ) = 2.13; RMSEA = 0 (90% CI: 0–0.116); SRMR = 0.04].

Following the preferred model, the pattern of results largely resembled that of the model for the full sample. A male participant of either group with a mean GDS score (group mean) at baseline would experience a drop in GDS score by  $-1.52$  points during the intervention ( $\beta_0 = -1.52$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 3.21$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.07$ ) and by  $-0.49$  during the first follow-up period ( $\beta_0 = -0.49$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 0.57$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.55$ ); both reductions were not significant. As in the total sample, there was a significant group difference in the effect of gender on change during the intervention ( $\Delta\chi^2_{S-B\text{-scaled}} = 6.54$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.01$ ).

More specifically, a significant effect of gender on  $\Delta\text{Dep}_{t_1t_0}$  was present in the IG ( $\beta_{\text{unstd}} = 3.92$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 14.79$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} < 0.001$ ; being a woman meant a reduction of  $-3.92$  GDS points more compared to being a man) but not in the CG ( $\beta_{\text{unstd}} = -0.62$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 0.23$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.63$ ). Together, this indicates the same treatment  $\times$  gender interaction effect as in the total sample, such that gender differences in the reduction of depressive symptoms during the intervention were only present in the IG and intervention effects were only present for female patients.

As shown in figure 2, these effects were more pronounced in the depressed subgroup than in the total sample. As in the total sample, baseline GDS score influenced  $\Delta\text{Dep}_{t_1t_0}$  similarly in both groups ( $\beta_{\text{unstd}} = -0.31$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 5.99$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.014$ ; the higher the baseline GDS score, the higher the reduction during the intervention). Concerning the follow-up period,  $\Delta\text{Dep}_{t_1t_0}$  influenced  $\Delta\text{Dep}_{t_2t_1}$ , indicating fluctuations in the GDS score ( $\beta_{\text{unstd}} = -0.45$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 8.57$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.003$ ; the higher the GDS score after the intervention, the lower it was after follow-up). Importantly, as in the total sample, these values did not differ between groups. In contrast to the results from the total sample, the baseline GDS score influenced also  $\Delta\text{Dep}_{t_2t_1}$  in the IG ( $\beta_{\text{unstd}} = -0.54$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 7.37$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.007$ ; the higher one started, the more reduced a score one had in the follow-up period), but it did not so in the CG ( $\beta_{\text{unstd}} = -0.08$ ;  $\Delta\chi^2_{S-B\text{-scaled}} = 0.27$ ,  $\Delta df = 1$ ,  $p_{\Delta\chi^2} = 0.60$ ).

In the CG, the preferred model explained 6.8% of the total variance of  $\Delta\text{Dep}_{t_1t_0}$ ; a model without gender explained 7.1%. In the IG, the preferred model explained 26.1% of the total variance of  $\Delta\text{Dep}_{t_1t_0}$ , and a model without gender explained 3.9%, showing that gender explained a large part of the variance in pre-post GDS change in the IG but not in the CG. For  $\Delta\text{Dep}_{t_2t_1}$ , the model explained 28.9% of the variance for the CG and 23.3% for the IG (without gender: 29.7 and 24.7%, respectively), indicating that gender did not add to explaining variance in the follow-up period.

## **Discussion**

LCSM of the longitudinal CORDIAL data set revealed that the best-fitting model was a decrease in depressive symptoms in the IG relative to the CG. Going beyond previous GLM-based analyses [18], the current analysis involved 4 assessment points, covering a period of 1 year and 3 months. Given the progressing character of AD,

this is an exceptionally long time window for therapy effects seldom reported in previous research.

The reduction in depressive symptoms was only observed in female patients in the IG. This finding is in accordance with previous studies reporting women to benefit more from CB therapy and supportive therapy [33]. However, these findings were not unequivocal [20, 34]. Clearly, further research is needed to investigate whether these gender differences would also apply to patients with AD. In the CORDIAL study, the gender effect is unlikely to result of differences in age or baseline levels of depressiveness, since these variables were statistically controlled for. Age was included as a control variable in post hoc analyses, which did not change the results. Possible explanations, as also discussed in the original publication [18], might be that specific parts of the therapy might be better suited to female patients. Alternatively, the benefit to male patients from the CB/CR intervention might have been smaller because their caregivers – preponderantly women, mostly their wives – might, even prior to therapy, have been more active e.g. in initiating social activities or ensuring a day structure based on their lifetime experience in family management. As a consequence, the subjectively perceived difference induced by the CB/CR intervention might have been smaller, resulting in more moderate differences in the GDS self-report measure. An additional depression scale filled in by the independent observers might have been valuable in this context.

As a second influencing factor, the baseline level of depression considerably influenced the degree of symptom reduction. This influence was independent of the group, but the total reduction in symptoms was generally smaller in the CG. It should be mentioned in this context that the presence of early AD, and not depression, was the inclusion criterion for the CORDIAL sample, resulting in a large variability of depressive symptoms. LCSM revealed that with every additional point above average on the GDS at baseline, the number of depressive symptoms was reduced by 0.33 GDS points at  $t_1$  in addition to the intercept reduction of 0.81 GDS points for an averagely depressed patient of either group. Combining the effects of the intervention and gender in the model for clinically depressed patients showed an undoubtedly clinically relevant reduction of 5.5 points on the GDS scale for an average female AD patient with a GDS score above the cutoff (fig. 2). As a possible limitation, the number of variables analyzed in our model was restricted due to the sample size. Further studies on this topic could, if involving larger samples or several mea-

sures per construct, incorporate factors that might further elucidate the intervention and gender effect found here.

In summary, the reported analyses may expand our knowledge about evaluation methods in two main respects. First, regarding clinical studies in general, our study may exemplarily show that LCSM is well suited to modeling data from clinical studies, especially those with high variability within groups, as is typical for the aged and for patients with age-associated cognitive disorders. In the future, modeling approaches such as LCSM may be extended to other research questions related to differential indication in the aged or in other subpopulations. For heart disease, there is evidence for a gender-specific efficacy of diagnosis and treatment [35]. Also, treatment effects may depend on prior interventions [36], or even be harmful for patients with certain cotherapies [37].

Second, our study may serve as an example for future practical applications of LCSM, as this method can provide concrete quantitative information about the effect of a given intervention on patients with specific features. As outlined above, an average reduction on the GDS scale can be forecast by the model, not only for the entire group but also for subgroups according to gender, IG, and baseline depressive symptoms. Concededly, this approach may enhance the complexity of diagnostics and treatment decisions, and translating this knowledge into daily clinical routines may still be challenging. Nevertheless, in accordance with the vision of individualized medicine in the 21st century, LCSM may represent a tool for tailoring interventions to patients in a truly evidence-based manner.

## Acknowledgments

Many thanks are due to the CORDIAL study board members Sylvia Egert, Lutz Froehlich, Herrmann-Josef Gertz, and Stefan Wagenpfeil as well as Angelika Thoenes-Otto for fruitful discussions and permission to analyze the data, and to Anima Pieper and Yasmina Giebler for assistance in manuscript preparation.

## Disclosure Statement

The authors have nothing to disclose.

## References

- 1 Knight BG, Sayegh P: Mental health and aging in the 21st century. *J Aging Soc Policy* 2011; 23:228–243.
- 2 Laidlaw K: A deficit in psychotherapeutic care for older people with depression and anxiety. *Gerontology* 2013;59:549–556.
- 3 Nelson EA, Dannefer D: Aged heterogeneity – fact or fiction – the fate of diversity in gerontological research. *Gerontologist* 1992;32:17–23.
- 4 Baltes MM, Maas I: Intraindividual change and interindividual differences in change: the Berlin Aging Study, BASE. *Z Gerontol Geriatr* 1999;32:18.
- 5 Schmider E, Ziegler M, Danay E, Beyer L, Bühner M: Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology (Gott)* 2010;6:147–151.
- 6 Lix LM, Keselman JC, Keselman HJ: Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance F test. *Rev Educ Res* 1996;66:579–619.
- 7 Tomarken AJ, Serlin RC: Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychol Bull* 1986;99:90–99.
- 8 Vickers AJ, Altman DG: Statistics notes – analysing controlled trials with baseline and follow-up measurements. *Br Med J* 2001;323: 1123–1124.
- 9 Kenny DA, Judd CM: Consequences of violating the independence assumption in analysis of variance. *Psychol Bull* 1986;99:422–431.
- 10 Vasey MW, Thayer JF: The continuing problem of false positives in repeated measures ANOVA in psychophysiology – a multivariate solution. *Psychophysiology* 1987;24:479–486.
- 11 Fava GA, Tomba E, Tossani E: Innovative trends in the design of therapeutic trials in psychopharmacology and psychotherapy. *Prog Neuropsychopharmacol Biol Psychiatry* 2013;40:306–311.
- 12 Duncan TE, Duncan SC, Strycker LA, Li F: An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications, ed 2. Mahwah, Erlbaum, 2006.
- 13 McArdle JJ: Latent variable modeling of differences and changes with longitudinal data. *Annu Rev Psychol* 2009;60:577–605.
- 14 Baltes PB, Smith J: New frontiers in the future of aging: from successful aging of the young old to the dilemma of the fourth age. *Gerontology* 2003;49:123–135.
- 15 Dannefer D: Cumulative advantage/disadvantage and the life course: cross-fertilizing age and social science theory. *J Gerontol* 2003; 58:327–337.
- 16 Ziegler M, Danay E, Heene M, Asendorpf J, Bühner M: Openness, fluid intelligence, and crystallized intelligence: toward an integrative model. *J Res Pers* 2012;46:173–183.
- 17 Moergeli H, Wittmann L, Schnyder U: Quality of life after traumatic injury: a latent trajectory modeling approach. *Psychother Psychosom* 2012;81:305–311.
- 18 Kurz A, Thöne-Otto A, Cramer B, Egert S, Frölich L, Gertz HJ, Kehl V, Wagenpfeil S, Werheid K: CORDIAL: cognitive rehabilitation and cognitive-behavioral treatment for early dementia in Alzheimer disease: a multicenter, randomized, controlled trial. *Alz Dis Assoc Dis* 2012;26:246–253.
- 19 Hy LX, Keller DM: Prevalence of AD among whites – a summary by levels of severity. *Neurology* 2000;55:198–204.
- 20 Baron S, Ulstein I, Werheid K: Psychosocial interventions in Alzheimer's disease and amnesic mild cognitive impairment: evidence for gender bias in clinical trials. *Aging Ment Health* 2015;19:290–305.
- 21 Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO: Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1983;17:37–49.
- 22 Guggel S, Birkner B: Validität und Reliabilität einer deutschen Version der Geriatrischen Depressionsskala. *Z Klin Neuropsychol* 1999;28:18–27.
- 23 Körner A, Lauritzen L, Abelskov K, Gulmann N, Marie Brodersen A, Wedervang-Jensen T, Marie Kjeldgaard K: The Geriatric Depression Scale and the Cornell Scale for Depression in Dementia. A validity study. *Nord J Psychiatry* 2006;60:360–364.
- 24 Debruynne H, Van Buggenhout M, Le Bastard N, Aries M, Audenaert K, De Deyn PP, Engelborghs S: Is the Geriatric Depression Scale a reliable screening tool for depressive symptoms in elderly patients with cognitive impairment? *Int J Geriatr Psychiatry* 2009;24: 556–562.
- 25 Muthén L, Muthén B: Mplus User's Guide. Los Angeles, Muthén & Muthén, 2008.
- 26 Kline RB: Principles and Practice of Structural Equation Modeling. New York, Guilford, 2011.
- 27 Chen FF: Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Modeling* 2007;14:464–504.
- 28 Satorra A, Bentler P: A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 2001;66:507–514.
- 29 Arbuckle JL: Full information estimation in the presence of incomplete data; in Marcoulides GA, Schumacker RE (eds): *Advanced Structural Equation Modeling*. Mahwah, Lawrence Erlbaum Associates, 1996, pp 243–277.
- 30 Schafer JL, Graham JW: Missing data: our view of the state of the art. *Psychol Methods* 2002;7:147–177.
- 31 Faul F, Erdfelder E, Buchner A, Lang AG: Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 2009;41:1149–1160.
- 32 Cohen J: *Statistical Power Analysis for the Behavioral Sciences*, ed 2. Hillsdale, Lawrence Erlbaum Associates, 1988.
- 33 Ogrodniczuk JS, Piper WE, Joyce AS, McCallum M: Effect of patient gender on outcome in two forms of short-term individual psychotherapy. *J Psychother Pract Res* 2001;10:69–78.
- 34 Clarkin JF, Levy KN: The influence of client variables on psychotherapy; in Lambert MJ (ed): *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. New York, Wiley, 2004, vol 5, pp 194–226.
- 35 Regitz-Zagrosek V, Oertelt-Prigione S, Seeland U, Hetzer R: Sex and gender differences in myocardial hypertrophy and heart failure. *Circ J* 2010;74:1265–1273.
- 36 Tomba E: Nowhere patients. *Psychother Psychosom* 2012;81:69–72.
- 37 Horwitz RI, Singer BH, Makuch RW, Viscoli CM: Can treatment that is helpful on average be harmful to some patients? *J Clin Epidemiol* 1996;49:395–400.